

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings of claims in the application. Applicant has submitted a new complete claim set showing any marked up claims with insertions indicated by underlining and deletions indicated by strikeouts and/or double bracketing.

1. (Currently Amended) A system for performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith and an external domain name server (DNS-B), the system comprising:

~~a first domain name server deployed on an Internet backbone (DNS-B); and~~
one of a plurality of load balancing domain name servers (DNS-LBs) deployed in a physical proximity from which the actual network latency of the clients may be measured, the DNS-LBs having stored therein IP address information of the multiple globally-dispersed servers to be load balanced, the DNS-LBs each sending mapping information to the DNS-B relating the DNS-LB's IP address to an IP address of the DNS-ISP to which the DNS-LB is in a physical proximity from which the actual network latency of the clients may be measured, the DNS-LBs determining performance characteristics of each of the multiple globally-dispersed servers, a DNS-LB receiving DNS lookup requests sent from its respective physically-proximate clients to the DNS-LB's corresponding DNS-ISP, the DNS lookup requests comprising respective hostnames of some of the globally-dispersed servers, the DNS-LB using its measurements of actual network latency to resolve the DNS lookup requests to respective IP addresses of the some of the globally-dispersed servers, where DNS lookup request's hostname can be resolved

to multiple of the IP addresses and the DNS-LB returns to the client the IP address that has lower network latency.

2. (Original) The system of claim 1, wherein the DNS-B stores the mapping information for the plurality of DNS-LBs to forward IP address queries to one of the DNS-LBs closest to the DNS-ISP from which the IP address query originated, and wherein the DNS-LB closest to the DNS-ISP returns the IP address to the DNS-ISP of the server having the best performance characteristics.

3. (Original) The system of claim 1, wherein the DNS-B stores the mapping information for the plurality of DNS-LBs to forward IP address queries to one of the DNS-LBs closest to the DNS-ISP from which the IP address query originated, and wherein the DNS-LB closest to the DNS-ISP returns the IP address of the DNS-LB to the DNS-ISP.

4. (Original) The system of claim 1, wherein the DNS-B provides its IP address information to the DNS-A to enable the DNS-A to forward IP address queries to the DNS-B.

5. (Original) The system of claim 4, wherein the DNS-B receives IP address information from the DNS-A for the servers to be load balanced.

6. (Original) The system of claim 1, wherein the DNS-LB is a client of the DNS-ISP.

7. (Original) The system of claim 1, further comprising a DNS-B deployed on each Internet backbone, and wherein each DNS-B contains the mapping information for all of the DNS-LBs stored therein.

8. (Original) The system of claim 1, wherein the DNS-LB transmits updated mapping information upon a change of an IP address of the DNS-ISP.

9. (Original) The system of claim 1, wherein each of the DNS-LBs transmit performance information of the servers to the DNS-B, and wherein the DNS-B utilizes the mapping information to determine the proper DNS-LB performance information to utilize to select the IP address of the server having the best performance characteristics to return to the DNS-ISP from which an IP address query originated.

10. (Currently Amended) A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith, the method comprising the steps of:

receiving IP address information from the DNS-A for the servers to be load balanced;

providing the IP address information to a plurality of load balancing domain name servers (DNS-LB);

receiving mapping information associating DNS-ISP IP address information to IP address information of a DNS-LB located in a physical proximity from which the actual network latency of the clients may be measured, the DNS-LB capable of determining server performance from a location physically proximate to the ISP's point of presence; and

referring DNS address inquiries from a DNS-ISP to a physically proximate DNS-LB in accordance with the mapping information, a DNS address inquiry comprising a hostname corresponding to multiple of the globally-dispersed servers, and wherein the DNS-LB selects one of the multiple servers according to the DNS-LB's determining of server performance and answers the DNS address inquiry by returning the IP address of the selected server.

11. (Original) A computer-readable medium having computer executable-instructions for performing the steps of claim 10.

12. (Currently Amended) A method of performing client-centric load balancing of multiple globally-dispersed servers, the servers being accessed by clients connecting through an ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith, the method comprising the steps of:

obtaining, by a load balancing domain name server (DNS-LB), IP address information for a DNS-ISP, the DNS-LB located in a physical proximity from which the actual network latency of the clients may be measured;

providing a mapping of an IP address of the DNS-LB to the IP address information of the DNS-ISP to an external domain name server;

receiving IP address information for the servers;

monitoring performance of the servers at the received IP addresses by the DNS-LB
transmitting communications to the IP addresses of the servers;

receiving at the DNS-LB a DNS name query that was sent from one of the clients to the DNS-ISP, the DNS name query querying for an IP address of a hostname that corresponds to multiple of the servers; and

providing at least one IP address for a server in response to a to the DNS name query selected by selecting the server, based on the monitoring step, from among the multiple servers that correspond to the hostname, and by returning the IP address for the selected server.

13. (Original) The method of claim 12, further comprising the steps of:

detecting a change in the DNS-ISP IP address; and

updating the mapping of the IP address of the DNS-LB to the IP address information of the DNS-ISP to the external domain name server.

14. (Original) The method of claim 12, further comprising the steps of
receiving selection criteria for the selection of an IP address;
receiving a name query from the DNS-ISP; and
wherein the step of providing at least one IP address for a server in response to a name query selected based on the monitoring step further comprises the step of providing at least one IP address for a server in response to a name query selected based on the monitoring step and on the selection criteria.

15. (Original) A computer-readable medium having computer-executable instructions for performing the steps of claim 12.

16-19. (Canceled)

20. (Currently Amended) A method of performing client-centric load balancing of multiple globally-dispersed content servers for handling content requests from clients, the servers being accessed by clients connecting through an Internet service provider's (ISP's) ~~at a point of~~ presence (POP), ~~each the~~ ISP having a domain name server (DNS-ISP), the servers further having an authoritative domain name server (DNS-A) associated therewith containing information regarding the IP addresses of the servers, the method comprising the steps of:

deploying a ~~first~~ plurality of load balancing domain name servers (DNS-LBs) in a physical proximity from which the actual network latency of the clients connecting to the ISP POPs may be measured;

~~deploying a second plurality of second level domain name servers (DNS-Bs) on the Internet backbones and regional providers;~~

communicating IP address information for ~~the~~ a plurality of second level domain name servers (DNS-Bs) to the DNS-As to enable the DNS-As to refer name queries to the DNS-Bs;

Microsoft Corporation
Application No.: 09/714,406
Filed: Nov 16, 2000

providing, by the DNS-LBs to the DNS-B, mapping information associating ~~an~~ IP addresses of the DNS-LBs ~~to the DNS-ISP to an~~ IP addresses of their corresponding DNS-ISPs to enable the DNS-B to refer name queries ~~from a~~ from DNS-ISPs to DNS-LBs ~~to the DNS-ISP~~; and communicating IP address information of the servers to the DNS-LBs;

monitoring, by the DNS-LBs, performance of the servers;

receiving at a DNS-LB a DNS name query that was sent from one of the clients to the DNS-ISP, the DNS name query querying for an IP address of a hostname that corresponds to multiple of the servers, the DNS name query having been sent from the client in response to the client starting a service request needing an IP address for the hostname; and

providing, by the DNS-LBs, in response ~~to a~~ to the DNS name query from the DNS-ISP, the IP address of a server by selecting the IP address from among the IP addresses of the multiple of the servers based on the step of monitoring.

21. (Cancelled)

22. (Cancelled)

23. (Currently Amended) A method for load balancing content servers, each of the content servers being associated with a same domain name, the method comprising:

receiving a DNS request to resolve the domain name from an ISP DNS server;

identifying at least one load balancing server from a group of load balancing servers, the identified load balancing server ~~situated at a physical proximity from which the actual measuring network latency from the load balancing servers to the content servers of a client connecting to the ISP DNS server may be measured; and~~

sending the IP address of the identified load balancing server to the ISP DNS server, the identified load balancing server configured to ~~determine~~ select, based on its measuring of network latency, at least one of the content servers ~~with certain characteristics relative to the~~

Microsoft Corporation
Application No.: 09/714,406
Filed: Nov 16, 2000

~~location~~ and to resolve the domain name with an IP address associated with the ~~determined~~
selected content server.

24. (Previously Presented) The method as recited in claim 23, wherein the certain characteristics include load level, availability, network latency, or network cost.

25. (Previously Presented) The method as recited in claim 23, wherein the identified load balancing server is situated closest to the ISP DNS server among the group of load balancing servers.

26. (Cancelled)

27. (Previously Presented) The system as recited in claim 26, wherein the certain characteristics include load level, availability, network latency, or network cost.

28. (New) A method comprising:

receiving at a load-balancing domain name service server (DNS-LB) a DNS lookup request received by and redirected from a domain name service server of an internet service provider (DNS-ISP), the request having been sent by a client of the DNS-ISP, the request containing a hostname corresponding to a plurality of IP addresses of servers serving content, where the request was forwarded by the DNS-ISP when the DNS-ISP determined that an IP address for the hostname was not cached at the DNS-ISP and obtained the IP address of the DNS-LB by issuing a DNS query for the hostname;

measuring network latency from the DNS-LB to the servers that correspond to the hostname in the request by repeatedly sending communications from the DNS-LB to the servers;

in response to receiving the redirected DNS lookup request of the client at the DNS-LB, selecting an IP address of one of the servers that correspond to the hostname, where the IP address is selected from among the servers based on the measuring of network latency to the servers; and

returning to the client the selected IP address.

29. (New) A method performed by a DNS server that provides DNS service for a plurality of clients, the method comprising:

receiving from a client a DNS lookup request requesting an IP address for a server having a hostname specified by the request;

in response to receiving the client DNS lookup request, determining that an IP address for the requested hostname is unavailable on the DNS server and in response issuing a DNS query for the hostname;

in response to issuing the DNS query for the hostname by the DNS server, receiving a referral to an authoritative DNS server (DNS-A) that corresponds to the hostname, the referral providing an IP address of a domain name service load-balancing server (DNS-LB) and causing the DNS server to query the DNS-LB for an IP address of the hostname in the client request;

in response to querying the DNS-LB, receiving an IP address from the DNS-LB, where the IP address corresponding to the hostname was selected by the DNS-LB based on network measurements obtained by repeated transmissions from the DNS-LB to IP addresses that correspond to the hostname; and

sending the IP address received from the DNS-LB to the client that sent the lookup request for the web server having the hostname specified by the request.